Truncated Gaussians as Tolerance Sets*

Fabio Cozman Eric Krotkov
Robotics Institute, Carnegie Mellon University
Pittsburgh, PA - 15213
e-mail: fgcozman@cs.cmu.edu

June 23, 1997

1 Introduction

This work presents a new class of statistical models that are well suited for several Robotics applications, such as object recognition or computer vision. Our approach deals with bounded data: measurements that are constrained to appear in a bounded region in the measurement space. The literature refers to the set where the measurement can appear as the *tolerance set* for the measurement.

To date, few statistical models for bounded variables are used in Artificial Intelligence. The most common model is the uniform distribution, but this approach has several drawbacks: summation of uniform variables does not yield a uniform variable and application of Bayes rule is computationally intensive [9]. Another approach is to use the Gaussian distribution and model bounds through an *ad hoc* selection mechanism [1, 2]. In short, even though bounds contain significant information, they have not yet received proper attention.

Our work uses a class of distributions in the truncated Gaussian family to model bounded data. Such models have been used in Statistics in order to implement selection mechanisms [7]. Instead, we propose truncated models not only as models of selected data but also as models for bounded noise, e.g., disturbances that appear in object recognition and computer vision [6]. We derive a complete set of novel algorithms for these models: calculation of moments, approximation methods for Bayes rule and summation and noise filtering. Overall, our results make the truncated Gaussian family an operational tool, more powerful than the uniform or the Gaussian distributions. Our analysis complements results scattered in the literature of Statistics, Information Theory and Control Theory. We contribute to Artificial Intelligence by indicating a proper model of bounded measurements and by deriving tractable algorithms to handle them.

2 The Truncated Gaussian Family

Our basic model is the elliptically truncated Gaussian family. A distribution in this family is proportional to a Gaussian inside an ellipsoid and is zero outside the ellipsoid. The truncated Gaussian model has been proposed in a variety of contexts in Statistics [7] as models of selection mechanisms. In this work we focus on the truncated Gaussian as a model for measurement noise, not so much as a model for selected measurements. The first explicit mention of this possibility in Robotics appears to be by Erdmann [4].

*This research is supported in part by NASA under Grant NAGW-1175. Fabio Cozman is supported under a scholarship from CNPq, Brazil.

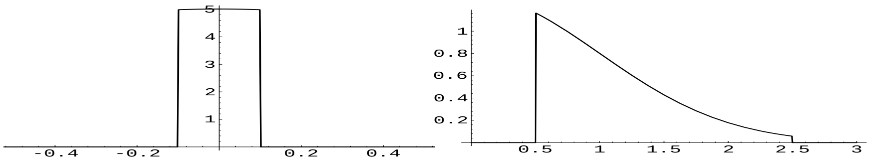

Figure 1: Distributions in the truncated Gaussian family: $N_{0,1,0.01}(0,1)$ (left) and $N_{1.5,1,1}(0,1)$ (right).

A truncated Gaussian distribution for an n-dimensional random vector x is referred to as $N_{\nu,M,k}(\mu,P)$; its mathematical expression is:

$$N_{\nu,M,k}(\mu,P) = \frac{(q(\mu,P,\nu,M,k))^{-1}}{|(2\pi)^n \det(P)|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^T P^{-1}(x-\mu)\right) I_{\{x:(x-\nu)^T M^{-1}(x-\nu)\leq k\}}(x), \qquad (1)$$

where $I(\cdot)$ is the indicator function and $q(\mu,P,\nu,M,k)$ is a normalizing constant. The set $\{x : (x-\nu)^T M^{-1}(x-\nu) \leq k\}$ defines an ellipsoid in n-dimensional space. Call k the *radius* of the distribution.

As a special case, note that if $\mu=\nu$ and $M=P$, then the normalizing constant depends only on k. In this case, $q(\mu,P,\mu,P,k) = Pr(\chi^2_n \leq k)$ (chi-square distribution). Due to the importance of this sub-family, we call it the *radially truncated Gaussian* family.

The truncated Gaussian is a suitable model for bounded data, even more powerful than the usual Gaussian since it can represent nearly flat and highly skewed distributions. Figure 1 illustrates this claim.

Suppose we know a random vector x has mean μ, covariance matrix Q, and $Pr(x)=0$ for all x outside $\{x : (x-\mu)^T Q^{-1}(x-\mu) \leq k\}$. In other words, possible values of x concentrate around the mean in a symmetric fashion, up to the distance k in the metric induced by Q. We can strengthen the parallel between truncated and unbounded Gaussians by proving that (all proofs can be found in [3]):

Theorem 1 *A maximum entropy distribution, when the expected value is μ, the covariance matrix is Q and the distribution is zero outside $\{x : (x-\mu)^T Q^{-1}(x-\mu) \leq k\}$, is a truncated Gaussian $N_{\mu,cQ,k'}(\mu,cQ)$, where $c = Pr(\chi^2_n \leq k)(Pr(\chi^2_{n+2} \leq k))^{-1}$ and $k' = kPr(\chi^2_{n+2} \leq k)(Pr(\chi^2_n \leq k))^{-1}$.*

2.1 Moment Generating Function, Mean and Variance of a Truncated Gaussian

The central problem in the characterization of a truncated Gaussian is the determination of $q(\mu,P,\nu,M,k)$. In this section, we consider a linear transformation on the initial random variable such that:

$$N_{\omega,D,k}(0,I) = \frac{(q(\omega,D,k))^{-1}}{(2\pi)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}z^T z\right) I_{\{z:(z-\omega)^T D^{-1}(z-\omega)\leq k\}}(z),$$

where D is a diagonal matrix, using a linear transformation: $z = \Phi^T \sqrt{\Lambda}^{-1} V^T (x-\mu)$, where: Λ is the eigenvalue matrix of P, V is the corresponding eigenvector matrix of P and Φ is the eigenvector matrix of $(\sqrt{\Lambda}^{-1} V^T) M (V \sqrt{\Lambda}^{-1})$. Details can be found in [7] or [3].

We work with the transformed variable z, since $q(\omega,D,k) = q(\mu,P,\nu,M,k)$. Then ($d_i$ the inverse of the ith element in the diagonal of D): $q(\omega,D,k) = Pr\left(\sum_{i=1}^n d_i(z_i-\omega_i)^2 \leq k\right)$. Kotz, Johnson and Boyd [7] give a numerical method for the evaluation of this integral based on its Laguerre expansion. We have:

$$q(\omega,D,k) = Pr(\chi^2_n \leq k/\beta) + \left(\frac{k}{2\beta}\right)^{n/2} e^{-\frac{k}{2\beta}} \sum_{j=1}^{\infty} c_j L_{j-1}(k/(2\beta)) \frac{(j-1)!}{\Gamma(n/2+j)}$$

where $L_r(y)=\sum_{i=0}^{r}\frac{(-y)^i}{i!(r-i)!}\frac{\Gamma(r+n/2+1)}{\Gamma(i+n/2+1)}$, $c_r=(2r)^{-1}\sum_{j=0}^{r-1}s_{r-j}c_j$, $c_o=1$ and $s_r=(-r/\beta)\sum_{j=1}^{n}\omega_i^2 d_i(1-d_i/\beta)^{r-1}+\sum_{j=1}^{n}(1-d_j/\beta)^r$.

Convergence is uniform for $\beta>\frac{\max_j(d_j)}{2}$. In general, large values of β yield slow convergence. If we truncate the evaluation of the series at $j=N$, the truncation error is always smaller than $2^{-N}\exp\left(k/(2\beta)+(n-1)\log 2+4\omega^T\omega\right)$.

The moment generating function of a truncated Gaussian $N_{\omega,D,k}(0,I)$ is $\phi(t)=\frac{q(\omega-t,D,k)}{q(\omega,D,k)}\exp\left(\frac{t^Tt}{2}\right)$. For a proof, refer to [3].

The mean and covariance can be obtained by successive differentiations of $\phi(t)$ [8]. Since Kotz, Johnson and Boyd recursions for $\phi(t)$ are uniformly convergent we can differentiate these expressions term by term. We use $\overline{\mu}$ for the mean and $\overline{P}$ for the covariance of a truncated Gaussian. Direct differentiation of the moment generating function yields:

Theorem 2 *We have:* $\overline{\mu}=k_o\sum_{j=1}^{\infty}k_jg_j$ *and* $\overline{P}=I-\overline{\mu}\overline{\mu}^T+k_o\sum_{j=1}^{\infty}k_jG_j$*, where (*$k_r$ *are scalars,* g_r *and* h_r *are vectors,* G_r *and* H_r *are matrices and* I *is an identity matrix):*

$$
\begin{aligned}
k_j &= \frac{(j-1)!L_{j-1}(k/(2\beta))}{\Gamma(n/2+j)} \quad , \quad k_o=\left(\frac{k}{2\beta}\right)^{n/2}\frac{e^{-\frac{k}{2\beta}}}{q(\omega,D,k)} \\
g_r &= (2r)^{-1}\sum_{j=0}^{r-1}\left(h_{r-j}c_j+s_{r-j}g_j\right) \quad , \quad g_0=0 \\
h_r &= \frac{2r}{\beta}\,\mathrm{diag}\,[d_1(1-d_1/\beta)^{r-1},\ldots,d_n(1-d_n/\beta)^{r-1}]\omega \\
G_r &= (2r)^{-1}\sum_{j=0}^{r-1}\left(c_jH_{r-j}+g_jh_{r-j}^T+h_{r-j}g_j^T+s_{r-j}G_j\right) \quad , \quad G_0=0 \\
H_r &= -\frac{2r}{\beta}\,\mathrm{diag}\,[d_1(1-d_1/\beta)^{r-1},\ldots,d_n(1-d_n/\beta)^{r-1}]
\end{aligned}
$$

For a distribution $N_{\mu,P,k}(\mu,P)$ in the radially truncated Gaussian family, the mean is μ and the covariance matrix is $\left(Pr(\chi^2_{n+2}\le k)(Pr(\chi^2_n\le k))^{-1}P\right)$ [7].

2.2 Linear Transformation and Summation of Truncated Gaussians

A non-singular linear transformation applied to a random vector with truncated Gaussian distribution produces another truncated Gaussian random vector, so if $x\sim N_{\nu,M,k}(\mu,P)$ and $y=Ax$ (where A is any non-singular square matrix) then $y\sim N_{A\nu,AMA^T,k}(A\mu,APA^T)$.

On the other hand, the sum of truncated Gaussian random variables is not a truncated Gaussian. We can derive some results for particular cases. Suppose $z=x+y$ where $x\sim N_{\nu_x,M_x,k_x}(\mu_x,P_x)$ and $y\sim N_{\nu_y,M_y,k_y}(\mu_y,P_y)$, x and y independent. Under these conditions:

Theorem 3 *The distribution of* z *has expected value* $\overline{\mu}_z=\overline{\mu}_x+\overline{\mu}_y$ *and covariance* $\overline{P}_z=\overline{P}_x+\overline{P}_y$*; the distribution of* z *is positive only inside the ellipsoid defined by* $\{z:(z-\nu_z)^TM_z^{-1}(z-\nu_z)\le 1\}$ *where*

$$\nu_z=\nu_x+\nu_y \qquad\qquad M_z=k_xM_x+k_yM_y$$

Therefore, a reasonable approximation to the distribution of z is $N_{\nu_z,M_z,1}(\overline{\mu}_z,\overline{P}_z)$.

A special but important case is represented by $\nu_x = \mu_x$, $M_x = P_x$, $\nu_y = \mu_y$, $M_y = P_y$, $k_x = k_y$. In this special case we can make statements about the distance between the approximation and the correct distribution. Call G_z the correct distribution for z; then:

Theorem 4 *For* $\mu_z = \mu_x + \mu_y$, $M_z = M_x + M_y$, $k_z = k_x = k_y$, $\sup_z |G_z - N_{\nu_z,M_z,k_z}(\nu_z, M_z)|$ *is* $\mathcal{O}(\exp(-k/2))$.

In this case, the indicated approximation is fairly good.

In the most general case, we can expect that the summation of truncated Gaussian can be well approximated by a Gaussian, by invoking the Central Limit Theorem. So we can expect a good approximation in terms of truncated Gaussians. This problem is discussed in [3], but no provably good approximation for all cases was found yet.

3 Inferences with the Truncated Gaussian

Inferences about a random variable are obtained by application of Bayes rule associated with a decision rule. Since the truncated Gaussian is not closed under multiplication, we should not expect to apply Bayes rule and obtain a truncated Gaussian distribution. Fortunately, we can find a good approximation for a situation relevant to practical applications. We consider a linear set-up, given by $z = Hx + \omega_z$, where x and z are vectors, A and H are matrices of appropriate dimensions, x is distributed as a truncated Gaussian $N_{\nu_x,M_x,k_x}(\mu_x, P_x)$; ω_z is distributed as a radially truncated Gaussian with zero mean $N_{0,P_z,k_z}(0, P_z)$. As a result, $z \sim N_{Hx,P_z,k_z}(Hx, P_z)$.

The central question is: *given that we observe the value of z, what can we say about x?*

there are two issues in finding the posterior distribution: obtaining the analytic expression for the density *and* obtaining the region of intersection between the positive regions of the prior and the likelihood. We solve these problems in turn.

3.1 Expression of the Posterior

The posterior distribution is proportional to: $\exp\left(-\frac{1}{2}(x-\mu_x)^T P_x^{-1}(x-\mu_x) + (z-Hx)^T P_z^{-1}(z-Hx)\right)$. Some manipulations reduce it to:

$$\frac{c}{|2\pi \det(Q)|^{n/2}} \exp\left(-\frac{1}{2}(x-\mu')^T Q^{-1}(x-\mu')\right) I_{A\cap B}(x), \tag{2}$$

where $Q = P_x - P_x H^T (H P_x H^T + P_z)^{-1} H P_x$ and $\mu' = Q(P_x^{-1}\mu_x + P_z^{-1}H^T z)$, and A and B are the regions where the prior and the likelihood are positive respectively (c is a normalizing constant).

3.2 Intersecting Ellipsoids: Fogel-Huang Algorithm

In order to approximate the posterior, an ellipsoidal approximation can be built so that the approximate posterior is always a truncated Gaussian. We indicate truncated ellipsoids by $N_{(\nu,M,k)}$ in this section.

We make some preliminary approximations:

- Suppose $\omega_z \sim N_{(0,W,1/a)}$ and $x \sim N_{(\mu,P,b)}$.
- Now take a linear transformation A such that $AWA^T = I$ (always possible since W must be positive definite). So we have $Az = AHx + A\omega_z$. Now consider $B = AH$, $y = Az$ and $\omega_y = A\omega_y$; we have $y = Bx + \omega_y$. and $\omega_y \sim N_{(0,I,1/a)}$.
- We know that $a\omega_y^T\omega_y \leq 1$; instead of trying to satisfy this inequality, we will satisfy a set of inequalities of the form: $a\omega_{yi}^2 \leq 1$ for $i = 1, \ldots, m$. By rearranging terms we have:

$$a(y_i - B_i x)^2 \leq 1 \text{ for } i = 1, \ldots, m. \tag{3}$$

  where y_i is the ith element of y and B_i is the ith row of B.

Now the problem is how to approximate the following set:

$$\Theta_m = \left\{x : \left[(x-\mu)^T P^{-1}(x-\mu) \leq b\right] \cap \left[\cap_1^m a(y_i - B_i x)^2 \leq 1\right]\right\}.$$

An approximation method was originally proposed by Fogel and Huang in the context of tolerance sets for ARMA processes. The algorithm finds the minimal enclosing ellipsoid to the intersection. Details can be found in [5] or [3].

The posterior distribution is a combined result of expression (2) and Fogel/Huang algorithm; we use $N_{\theta_m,P_m,b_m}(\mu', Q)$ as the posterior.

Example 1 *Consider the situation (where x is unknown):*

$$z = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix} x + \omega \qquad \omega \sim N_{0,I,1}(0, I) \qquad x \sim N_{a,B,1/4}(a, B) \qquad a = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad B = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \qquad z = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

Figure 2.a shows the initial ellipsoids of the problem. The region of possible values of x is the intersection of ellipsoids. Figure 2.b shows the strips generated at expression (3). The algorithm is applied and results are shown in figure 2.c. The largest hatched ellipse is the result of incorporating $z_1 = 2$; the smallest hatched ellipse is the final result after incorporation of $z_2 = 3$. The approximate posterior is $N_{c,D,e}(f, G)$ where $e = 0.257$ and:

$$c = \begin{bmatrix} 0.363 \\ 0.547 \end{bmatrix} \qquad D = \begin{bmatrix} 0.4488 & -0.3272 \\ -0.3272 & 0.7555 \end{bmatrix} \qquad f = \begin{bmatrix} 0.321 \\ 0.7418 \end{bmatrix} \qquad G = \begin{bmatrix} 0.091 & -0.0867 \\ -0.0867 & 1.8857 \end{bmatrix}. \square$$

The approximations found by the algorithm are correct in that the intersections are always interior to the successive approximations. Consider a different approximation strategy in the spirit of [1] or [2]. We would pretend x and ω to be distributed as unbounded Gaussians (with same mean and variance). Now we find a crucial problem: how to combine the fact that x has an elliptic region with radius $1/4$ and ω has an elliptic region with radius 1? By using Bayes rule in the unbounded Gaussians $N(0, I)$ and $N(a, B)$, we obtain an unbounded Gaussian $N(f, G)$. The hatched ellipses in figure 2.d shows the results of using these values of mean and variance with radius $1/4$, $5/8$ and 1. All ellipses fail to cover the region where x is known to lie. The case when the radius is $1/4$ is extreme: almost all values of x inferred from this posterior are inconsistent with prior expectations! Our methods have the advantage that, although approximate, they never assign probability zero to a possible value of the underlying unknowns.

3.3 Decision Rules

Two common estimates can be obtained from the posterior: the posterior mean and the maximum a posteriori. The posterior mean can be approximated from the results of last section. We discuss the maximum a posteriori here.

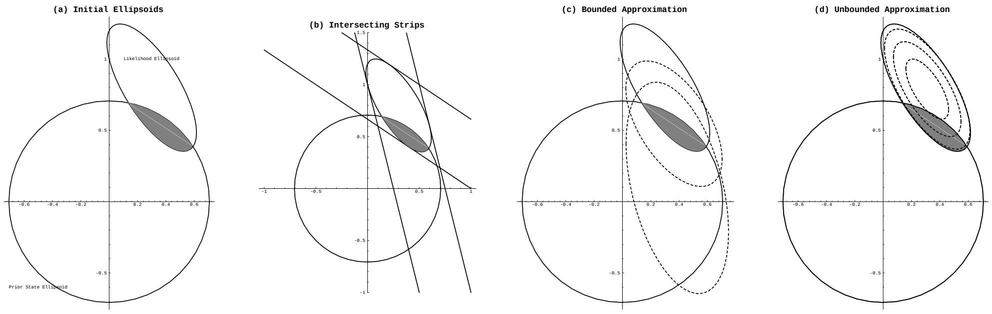

Figure 2: Example of Inference Algorithm for Bounded Distributions

Consider $x \sim N_{\nu,M,k}(\mu, P)$. We take the estimate $\hat{x} = \arg\max_x N_{\nu,M,k}(\mu, P)$. The optimization problem is simplified if we perform a double diagonalization. Consequently, we look for $\hat{w} = \arg\max_w N_{\xi,D,k}(0, I)$. There are two possible cases:

- If $\xi^T D^{-1} \xi \leq k$ then 0 is inside the defining ellipsoid. Since the Gaussian is unimodal, the maximum a posteriori is $\hat{w} = 0$ (it must be transformed back to x).
- Otherwise, the maximum occurs on the boundary of the defining ellipsoid. So we obtain a standard optimization problem: find the minimum of a quadratic function subject to a quadratic constraint. Solution by Lagrange multipliers is straightforward; details can be found in [3].

4 Filtering with the Truncated Gaussian

Consider first the simple linear system: $x_{i+1} = Ax_i$, $z_{i+1} = Bx_{i+1} + \omega_i$. We assume x_o distributed as a truncated Gaussian and ω distributed as zero mean radially truncated Gaussian. A and B are matrices of appropriate dimensions.

Given this set-up, we are interested in obtaining $p(x_{i+1}|z_1, \ldots, z_{i+1})$. Since $p(x_{i+1}|z_1, \ldots, z_{i+1}) \propto p(z_{i+1}|x_{i+1}) \times p(x_{i+1}|z_1, \ldots, z_i)$, the filtering scheme has the following structure: first, propagate forward the uncertainty in x_i; then use the system model and the Fogel/Huang algorithm in order to fuse the incoming information z_{i+1} with x_i. If necessary, the best estimate of x can be obtained at any time.

Extension of the filter above to a noisy state problem is straightforward, the result being very similar to Kalman filtering. Consider the system model: $x_{i+1} = Ax_i + \omega_{x_i}$, $z_{i+1} = x_{i+1} + \omega_{z_i}$. In this case we must obtain $p(x_{i+1}|z_1, \ldots, z_i)$ by calculating $p(Ax_i|z_1, \ldots, z_i)$ and combining it with $p(\omega_{z_i})$. Having done this, we can propagate forward the uncertainty in x_i (by properly approximating the summation of Ax_i and ω_{z_i}) and fuse it with the incoming information (through the Fogel/Huang algorithm).

5 Using the Truncated Gaussian: Disparity Constraint

In this section we briefly analyze a practical problem in Computer Vision using truncated Gaussians. The example will illustrate the power of the method.

Consider two cameras perfectly aligned so that all epipolar lines are parallel. Suppose a feature is detected in one camera at pixel x_2. The correspondence to the feature must lie in the epipolar line in the other camera in some pixel x_1. How much of the epipolar line should we explore in order to find the correspondence x_1? A powerful constraint that can be used is the fact that the point x_1 cannot be arbitrarily far from the given value of x_2. This constraint has not been justified or analyzed with the help of a statistical model. The validity of the constraint has been justified with biological analogies. But if we exploit the fact that the depth of field of any lens constrains depth, we can easily derive a statistical distribution for the disparity. In [3], we derive a truncated Gaussian distribution for the disparity, based on a truncated Gaussian distribution on the depth. The analysis incorporates the known facts of the problem without any analogies to biology.

6 Conclusion

The central idea of this research is: measurements that are bounded must be properly modeled in order to be consistently used. Our proposal is to use a family of statistical models (the truncated Gaussian family) that captures measurement boundedness. We offered a comprehensive analysis of estimation aspects for the truncated Gaussian: algorithms for information handling, updating and filtering. An open problem is to find good approximations for summations of truncated Gaussians.

Algorithms here developed complement existing work in Statistics in the area of selection mechanisms for data [7]. But our proposal goes beyond that, since we emphasize that many disturbances are, fundamentally, bounded. We propose truncated Gaussians as a replacement to several concepts generally used to deal with bounded data: the concept of tolerance sets (where the only source of information is the set of possible measurements) and the concept of validation gates (where *ad hoc* tests are used to simulated bounded distributions [1]). This work presents all the necessary tools to make the truncated Gaussian approach tractable and superior to these other approaches.

Acknowledgement

We gratefully acknowledge Ofer Barkai for his help in the proof of theorem 4. This paper benefited from several suggestions made by Mike Erdmann in the reading of a first draft.

References

[1] Y. Bar-Shalom and T. E. Fortmann, *Tracking and Data Association*, Academic Press, Boston, 1988.

[2] I. J. Cox, *A Review of Statistical Data Association Techniques for Motion Correspondence*, International Journal of Computer Vision, 10(1): 53-66, 1993.

[3] F. G. Cozman and E. Krotkov, *Truncated Gaussians as Tolerance Sets*, Technical Report CMU-RI-TR 94-35.

[4] M. Erdmann, *Randomization in Robot Tasks*, International Journal of Robotics Research, 11(5):399-436, October 1992.

[5] E. Fogel and Y. F. Huang, On the Value of Information in System Identification - Bounded Noise Case, *Automatica*, 18(2):229-238, 1982.

[6] W. E. L. Grimson, *Object Recognition by Computer: The Role of Geometric Constraints*, MIT Press, Cambridge, Mass, 1990.

[7] N. L. Johnson and S. Kotz, *Distributions in Statistics: Continuous Multivariate Distributions*, John Wiley and Sons, Inc., New York, 1972.

[8] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 3rd edition, McGraw-Hill, New York, 1991.

[9] L. D. Servi and Y. C. Ho, Recursive Estimation in the Presence of Uniformly Distributed Measurement Noise, *IEEE Trans. on Automatic Control*, AC-26(2):563-564, 1981.